

January 6, 2006

SRS Policy on Matching Data to Restricted-Use Data Sets

SRS occasionally receives requests to match SRS-collected survey data with administrative or other data as a potential research tool. SRS must review each such request on a case-by-case basis to determine the feasibility, merit and appropriateness of the activity. Any matching with SRS data will be conducted by SRS contractors (or, with SRS approval, by a sponsoring agency for the survey or the sponsoring agency's approved contractor) and the requester must pay the costs of creating and documenting the matched data set. SRS reserves the right to require "masking" of data items or use only at a secure site if the non-SRS data comes from a data set with identifiers readily available to the requesting researcher. Regardless of the incoming data source, the resulting matched data set, which will not contain any personal identifiers (i.e., names, Social Security numbers), will become an SRS data set, must be licensed (or may only be available for use at a secure site), will be governed by all regulations pertaining to such data sets, and may be made available to other researchers under an approved license.

Proposal for creating a matched data set

SRS seeks to create the most useful data sets possible and respond to the needs of the user community, but must balance the merits of each matching request against the need to protect the confidentiality of the SRS collected data, the potential impact on other SRS activities and use of SRS resources/contractors. Anyone wishing to request a matching of SRS data with other data must submit a proposal to SRS, which must contain:

- a description of the data sets to be matched,
- the research value of doing the matching,
- how the matched data set will be utilized,
- a description of the characteristics and quality of the non-SRS data that will be involved in the matching,
- a discussion of the variables and procedures necessary to do the matching, and
- the anticipated results of the matching (positive match rate, false positive rate, proportion of records that can be matched, etc).

SRS will review the feasibility, appropriateness, and value of undertaking a proposed match, including:

1. Legal issues, potential conflicts of interest for the researcher, and unacceptable disclosure risks that the proposed matched data set may present
 - A. Risk of disclosure—a matched data set may substantially increase the risk of disclosure of the respondents' identity.
 - B. To avoid excessive risk, "masking" of the data may be necessary but excessive masking reduces data utility.
 - C. Whether the data set can be licensed or can only be used at an NSF, NSF contractor, or survey sponsor site
2. Assessments of costs and benefits of the proposed match
 - A. Quality of the resulting data set
 - 1) Description of matching procedures
 - 2) Identification of issues relevant to the procedures and variables being used
 - 3) Anticipated results of the matching (positive match rate, false positive rate, proportion of records that can be matched, etc).
 - B. Time constraints in making the match
 - 1) Availability of the data sets to be matched
 - 2) Time schedules of the researcher and of SRS and its contractor
 - C. Gains from making the match
 - 1) Merit of the proposed research using the matched data set
 - 2) Potential for other applications/uses of the resulting data set
 - D. Potential costs of doing the match
 - 1) Increased demands upon SRS staff
 - 2) Additional demands upon contractor resources, that could reduce time devoted to SRS contract activities

Documentation of a matched data set

If SRS approves the proposed match and the requester executes an appropriate contract with the SRS contractor (or an approved agreement with a sponsoring agency), then the contractor conducting the match will prepare documentation of the match that will contain:

- an assessment of the data quality/characteristics of the non-SRS data used in the match,
- a description of the matching procedures, and
- an assessment of the quality of matching.

Review of the “new” data

SRS shall review *the data set for disclosure issues and provide written approval before any data are released to the researcher.*

Using a matched data set

A matched data set may be only used under license and/or at a secure site. The researcher must clearly understand that the matched data are not the property of the researcher - nor are any abstracts (e. g. subsets, recodes, constructed variables, etc.) of such data. As with all NSF licensed data, a researcher may not match the licensed data with other data, and in the case of a matched data set, that includes a prohibition to match the matched data set back to the original non-NSF data utilized in matching. The researcher must accept and abide by the following additional license conditions for the matched data set beyond the regular license conditions:

1. The researcher shall indicate on all products resulting from the use of the matched data set an NSF quality disclaimer – “NSF does not endorse the non-NSF data utilized in this report, does not assume responsibility for the accuracy of the non-NSF data, and does not necessarily endorse the research methodology used in the report.”
2. At the conclusion of the license, the researcher must return all data, the matched data as well, to SRS. No trace of such data may remain on storage media or hard drives of the researcher.

At the conclusion of the project, SRS will store the matched data, the documentation (especially data quality information), and related data in electronic form as may be necessary for the researcher to comply with any grant requirements for replicability of findings.

NSF may allow access to the matched data by other researchers under license or at a secure site provided such researchers agree to all the conditions and demonstrate a need for the merged data set.

This policy regarding matching to NSF restricted-use data sets is provisional, and is subject to change based on experience in implementing the policy.